Assessing GEE Models with Longitudinal Ordinal Data by Global Odds Ratio

LIN, KUO-CHIN
Graduate Institute of Business and Management
Tainan University of Technology
529 Chung-Cheng Road, Tainan 71002, Taiwan
Corresponding Author: t20053@mail.tut.edu.tw

CHEN, YI-JU
Department of Statistics, Tamkang University
151 Ying-Chuan Road, Taipei 25137, Taiwan
ychen@stat.tku.edu.tw

Abstract

Longitudinal studies are commonly occurred in the social and biomedical sciences. The overall test for model adequacy is an essential issue in longitudinal categorical data analysis. Two goodness-of-fit tests are proposed for the generalized estimating equations (GEE) models with longitudinal ordinal data in terms of Pearson chi-squared and unweighted sum of square residual using the global odds ratio as a measure of association. Our method can be regarded as an extension of the method proposed by Lin (Computational Statistics and Data Analysis 2010; 54:1872-1880) considering four major variants of working correlation structures. For large samples, the distributions of the two test statistics are approximated by the standard normal distribution based on the asymptotic means and variances. In simulation studies, the type I error rate and the power performance comparison between the proposed tests and the current tests are presented for various sample sizes. The application of the proposed tests is illustrated by a real data set.

Keywords: GEE, Global odds ratio, Goodness-of-fit, Longitudinal ordinal data.

1 Introduction

Longitudinal data involving a series of ordinal responses from each subject are frequently occurred in the social and biomedical sciences. Longitudinal ordinal data can be fitted by generalized estimating equations (GEE) models proposed by Liang and Zeger [1], or generalized linear mixed models (GLMMs) discussed by Breslow and Clayton [2]. Williamson *et al.* [3] developed marginal means in cumulative logit models and cumulative probit models based on a global odds ratio as the measure of association. Williamson *et al.* [4] also provided a SAS program, GEEGOR, for the analysis of correlated categorical responses. Heagerty and Zeger [5] adopted a proportional odds model with marginal cumulative probabilities using the GEE approach for clustered ordinal data. Yu and Yuan [6] extended program GEEGOR to program GGOREX for the analysis of longitudinal ordinal data.

Prior to making inferences on model parameters, the assessment of model fit is a critical issue. Many goodness-of-fit tests of regression models with longitudinal categorical responses have been discussed. Pan [7] considered two goodness-of-fit tests based on residuals, Pearson chi-squared and unweighted sum of residual squares, for GEE models with correlated binary data. Lin *et al.* [8] proposed a goodness-of-fit test for GEE models with longitudinal binary data using nonparametric smoothing. In this article, we propose the goodness-of-fit tests of regression models with longitudinal ordinal data utilizing the global odds ratio as a measure of association, which can be regarded as a generalization of Lin's methods [9] based on residual analysis and four major variants of working correlation structures.

The familiar tactics for analyzing ordinal responses are proportional odds models, adjacent category models and continuation ratio models referred by McCullagh [10], Clogg and Shihadeh [11], and Agresti [12]. Among of them, the cumulative logit model with proportional odds assumption is the most popular approach for analyzing ordinal data. The main aim of this article is to develop two goodness-of-fit tests of the proportional odds model for longitudinal ordinal data. In Section 2, the proportional odds models as well as the global odds ratio computational techniques are briefly introduced, and two goodness-of-fit tests for proportional odds models based on residual analysis using the global odds ratio are proposed. In Section 3, simulation studies are conducted for exploring the type I error rate and the power performance of the proposed tests. In addition, an example is employed to illustrate the proposed methods. Finally, it concludes with discussion and future research.

2 Goodness-of-fit Test Statistics

Let Y_{it} represent ordinal responses with $(J+1)$ categories, and $\mathbf{x}_{it}$ represent p-dimensional covariate vectors for subject i at occasion t for $i = 1, \ldots, n$ and $t = 1, \ldots, n_i$. The covariate vector $\mathbf{x}_{it}$ can be discrete or continuous. For simplicity, we assume equal occasions, $n_i \equiv T$. Denote $\mathbf{y}_{it}$ as a vector of J indicator variables, where $\mathbf{y}_{it} = (y_{it}^{(1)}, \ldots, y_{it}^{(J)})'$ with $y_{it}^{(j)} = 1$ if response $Y_{it} = j$ and 0 else. Let $\boldsymbol{\pi}_{it}$ and $\eta_{it}^{(j)}$ be the vector of marginal probabilities and the marginal cumulative probabilities, respectively, where $\boldsymbol{\pi}_{it} = (\pi_{it}^{(1)}, \ldots, \pi_{it}^{(J)})'$ with $\pi_{it}^{(j)} = P(Y_{it} = j|\mathbf{x}_{it}) = P(y_{it}^{(j)} = 1|\mathbf{x}_{it})$, and $\eta_{it}^{(j)} = P(Y_{it} \le j \,|\, \mathbf{x}_{it}) = \sum_{k=1}^{j} \pi_{it}^{(k)}$.

2.1 The Proportional Odds Model

The cumulative logit model with proportional odds assumption for describing the dependence of Y_{it} on $\mathbf{x}_{it}$ is given by $\text{logit}(\eta_{it}^{(j)}) = \log(\eta_{it}^{(j)}/(1-\eta_{it}^{(j)})) = \lambda_j + \mathbf{x}_{it}'\boldsymbol{\beta} = \xi_{it}^{(j)}$ for $j = 1, \ldots, J$, where the intercepts $\lambda_1, \ldots, \lambda_J$ satisfy $\lambda_1 \le \ldots \le \lambda_J$, $\boldsymbol{\beta}$ is the vector of regression coefficients with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and $\xi_{it}^{(j)}$ is the jth element of a J-dimensional linear predictor $\boldsymbol{\xi}_{it} = (\xi_{it}^{(1)}, \ldots, \xi_{it}^{(J)})'$. Owing to $\pi_{it}^{(1)} = \eta_{it}^{(1)}$ and $\pi_{it}^{(j)} = \exp(\xi_{it}^{(j)})(1+\exp(\xi_{it}^{(j)}))^{-1} - \exp(\xi_{it}^{(j-1)})(1+\exp(\xi_{it}^{(j-1)}))^{-1}$ for $j = 2, \cdots, J$, the linear predictor $\boldsymbol{\xi}_{it}$ can be rewritten as $\boldsymbol{\xi}_{it} = \mathbf{Z}_{it}'\boldsymbol{\theta}$ with the parameter vector $\boldsymbol{\theta} = (\lambda_1, \ldots, \lambda_J, \boldsymbol{\beta}')'$ and the $J \times (J+p)$ design matrix,

$$\mathbf{Z}_{it}' = \begin{bmatrix} 1 & & & \mathbf{x}_{it}' \\ & \ddots & & \vdots \\ & & 1 & \mathbf{x}_{it}' \end{bmatrix}.$$

A J-dimensional link function g connects $\boldsymbol{\pi}_{it}$ and the linear predictor $\mathbf{Z}_{it}'\boldsymbol{\theta}$ as $\boldsymbol{\pi}_{it} = g^{-1}(\mathbf{Z}_{it}'\boldsymbol{\theta})$.

Denote the responses, the marginal probabilities as well as the design matrix for subject i as $\mathbf{Y}_i = (\mathbf{y}_{i1}', \ldots, \mathbf{y}_{iT}')'$, $\boldsymbol{\pi}_i = (\boldsymbol{\pi}_{i1}', \ldots, \boldsymbol{\pi}_{iT}')'$, and $\mathbf{Z}_i = (\mathbf{Z}_{i1}, \ldots, \mathbf{Z}_{iT})'_{TJ\times(J+p)}$, respectively. The multivariate generalized estimating equations proposed by Lipsitz *et al.* [13] and Liang and Zeger [1] for estimating $\boldsymbol{\theta}$ is the solution to

$$\mathbf{v}_1(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\pi}_i(\boldsymbol{\theta})) = 0, \qquad (1)$$

where $\mathbf{D}_i = \partial\boldsymbol{\pi}_i/\partial\boldsymbol{\theta} = (\mathbf{D}_{i1}', \ldots, \mathbf{D}_{iT}')'$ with the jth row vector of $\mathbf{D}_{it}$ expressed by $(\partial\pi_{it}^{(j)}/\partial\lambda_1, \ldots, \partial\pi_{it}^{(j)}/\partial\beta_p)$ for $j = 1, \ldots, J$; $t = 1, \ldots, T$, and $\mathbf{V}_i = \text{Var}(\mathbf{Y}_i)$.

2.2 The Global Odds Ratio

Williamson *et al.* [3] proposed a moment method for the analysis of longitudinal ordinal responses, analogous to the methods of Prentice [14] and Lipsitz *et al.* [15], using two sets of generalized estimating equations: the marginal distribution as equation (1) and the global odds ratio for joint association between responses. The variance of $\mathbf{Y}_i$ is expressed by $\mathrm{Var}(\mathbf{Y}_i) \approx \mathbf{V}_i(\boldsymbol{\theta}, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is a vector of association parameters for describing the global odds ratio. The "working" covariance matrix of $\mathbf{Y}_i$,

$$\mathbf{V}_i(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} & \cdots & \mathbf{V}_{i1T} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} & \cdots & \mathbf{V}_{i2T} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{V}_{iT1} & \mathbf{V}_{iT2} & \cdots & \mathbf{V}_{iTT} \end{bmatrix}_{TJ\times TJ},$$

is a block matrix, where the diagonal block matrices, $\mathbf{V}_{itt} = \mathrm{Diag}(\boldsymbol{\pi}_{it}) - \boldsymbol{\pi}_{it}\,\boldsymbol{\pi}'_{it}$, describe the structural covariance matrices at a given point of time for $t = 1, \ldots, T$, and the off-diagonal block matrices $\mathbf{V}_{ist}$ have the elements of $\mathrm{cov}(y_{is}^{(j)}, y_{it}^{(k)}) = \mathrm{E}(y_{is}^{(j)} y_{it}^{(k)}) - \mathrm{E}(y_{is}^{(j)})\mathrm{E}(y_{it}^{(k)}) = \omega_{ijk}(s,t) - \pi_{is}^{(j)}\pi_{it}^{(k)}$ to express the longitudinal correlation matrices between any two time-points for $s, t = 1, \ldots, T$; $j, k = 1, \ldots, J$.

A global odds ratio is an extension of the simple odds ratio for a 2×2 contingency table. It is defined as the odds ratio of a 2×2 contingency table when the adjacent rows and columns of a $(J+1) \times (J+1)$ contingency table are collapsed into a 2×2 table. As such, a $(J+1) \times (J+1)$ table has J^2 global odds ratios. The global odds ratio for subject i with responses $Y_{is} = j$ and $Y_{it} = k$ is defined as

$$\Psi_{ijk}(s,t) = \frac{F_{ijk}(s,t)[1 - \eta_{is}^{(j)} - \eta_{it}^{(k)} + F_{ijk}(s,t)]}{[\eta_{is}^{(j)} - F_{ijk}(s,t)][\eta_{it}^{(k)} - F_{ijk}(s,t)]}, \tag{2}$$

where $F_{ijk}(s,t) = P(Y_{is} \le j, Y_{it} \le k)$ for $i = 1, \ldots, n$, j, $k = 1, \ldots, J$, and s, $t = 1, \ldots, T$. Plackett [16] showed that the joint distribution of $F_{ijk}(s,t)$ can be solved in terms of $\Psi_{ijk}(s,t)$ and the marginal cumulative probabilities:

$$\begin{aligned} F_{ijk}(s,t) &= \frac{[1 + (\eta_{is}^{(j)} + \eta_{it}^{(k)})(\Psi_{ijk}(s,t) - 1) - B_{ijk}(s,t)]}{2\,(\Psi_{ijk}(s,t) - 1)} && \text{if } \Psi_{ijk}(s,t) \ne 1, \\ &= \eta_{is}^{(j)}\,\eta_{it}^{(k)} && \text{if } \Psi_{ijk}(s,t) = 1, \end{aligned} \tag{3}$$

where $B_{ijk}(s,t) = \{[1 + (\eta_{is}^{(j)} + \eta_{it}^{(k)})(\Psi_{ijk}(s,t) - 1)]^2 + 4\Psi_{ijk}(s,t)(1 - \Psi_{ijk}(s,t))\eta_{is}^{(j)}\,\eta_{it}^{(k)}\}^{1/2}$ discussed by Williamson *et al.* [3]. The joint cell probabilities $\omega_{ijk}(s,t) = P(Y_{is} = j, Y_{it} = k) = F_{ijk}(s,t) - F_{ij(k-1)}(s,t) - F_{i(j-1)k}(s,t) + F_{i(j-1)(k-1)}(s,t)$ can be also determined by $\Psi_{ijk}(s,t)$, $\eta_{is}^{(j)}$ and $\eta_{it}^{(k)}$. A general model based on the global odds ratio is given by

$$log\Psi_{ijk}(s,t)(\boldsymbol{\alpha}) = \Delta + \Delta_j + \Delta_k + \Delta_{jk} + \boldsymbol{\gamma}'\mathbf{X}_{\mathrm{corr}_i}(s,t),$$

where $\boldsymbol{\alpha} = (\Delta, \Delta_j, \Delta_k, \Delta_{jk}, \boldsymbol{\gamma}')'$ is a vector of association parameters describing the global odds ratio, Δ, Δ_j, Δ_k, Δ_{jk} are an intercept term, the jth row effect, the kth column effect as well as the interaction effect with satisfying the constraints $\Delta_{J+1} = \Delta_{j(J+1)} = \Delta_{(J+1)j} = 0$ for $j = 1, \ldots, J$, respectively, and the parameters $\boldsymbol{\gamma}'$ corresponds to the covariates $\mathbf{X}_{\mathrm{corr}_i}(s,t)$.

Define $\mathbf{U}_i$ as a $(T(T-1)((J+1)^2 - 1)/2) \times 1$ random vector, $\mathbf{U}_i = (U_{i11}(1,2), \ldots, U_{i(J+1)J}(T-1,T))'$, and its expectation as $\mathrm{E}(\mathbf{U}_i) = \boldsymbol{\omega}_i = \boldsymbol{\omega}_i(\boldsymbol{\theta}, \boldsymbol{\alpha}) = (\omega_{i11}(1,2), \ldots, \omega_{i(J+1)J}(T-1,T))'$, where $U_{ijk}(s,t) = y_{is}^{(j)}\,y_{it}^{(k)}$. A set of generalized estimating equations is used for the estimation of unknown parameters $\boldsymbol{\alpha}$ as follows:

$$\mathbf{v}_2(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n} \mathbf{C}'_i\,\mathbf{W}_i^{-1}(\mathbf{U}_i - \boldsymbol{\omega}_i(\boldsymbol{\theta}, \boldsymbol{\alpha})) = 0, \tag{4}$$

where $\mathbf{C}_i = \partial\boldsymbol{\omega}_i/\partial\boldsymbol{\alpha}$ and $\mathbf{W}_i = \mathrm{Var}(\mathbf{U}_i)$. By equations (1) and (4), the estimates of $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}})$ are computed by an iterative algorithm:

$$\hat{\boldsymbol{\theta}}^{(l+1)} = \hat{\boldsymbol{\theta}}^{(l)} - \frac{\sum_{i=1}^{n} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} (\hat{\mathbf{Y}}_i - \boldsymbol{\pi}_i(\hat{\boldsymbol{\theta}}^{(l)}))}{\sum_{i=1}^{n} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i} \text{ and } \hat{\boldsymbol{\alpha}}^{(l+1)} = \hat{\boldsymbol{\alpha}}^{(l)} - \frac{\sum_{i=1}^{n} \hat{\mathbf{C}}_i' \hat{\mathbf{W}}_i^{-1} (\hat{\mathbf{U}}_i - \omega_i(\hat{\boldsymbol{\theta}}^{(l+1)}, \hat{\boldsymbol{\alpha}}^{(l)}))}{\sum_{i=1}^{n} \hat{\mathbf{C}}_i' \hat{\mathbf{W}}_i^{-1} \hat{\mathbf{C}}_i},$$

where l denotes the number of iterations.

2.3 Proposed Tests

Two proposed goodness-of-fit test statistics for assessing the proportional odds model with longitudinal ordinal data, Pearson chi-squared statistic and unweighted sum of residual squares statistic, are generalizations of the test statistics developed by Pan [7] for the logistic regression model with correlated binary data. The proposed test statistics are expressed by

$$G_p = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1}^{J} \frac{(y_{it}^{(j)} - \hat{\pi}_{it}^{(j)})^2}{\hat{\pi}_{it}^{(j)}(1 - \hat{\pi}_{it}^{(j)})} \text{ and } G_u = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{j=1}^{J} (y_{it}^{(j)} - \hat{\pi}_{it}^{(j)})^2, \tag{5}$$

where G_p is the Pearson chi-squared statistic and G_u is the unweighted sum of square residuals statistic.

The approximate expectations and variances of G_p and G_u derived by Lin *et al.* [17] , respectively, are given by $\widehat{\mathrm{E}(G_p)} = nTJ$, $\widehat{\mathrm{Var}(G_p)} = (1-2\hat{\boldsymbol{\pi}})'\hat{\mathbf{A}}^{-1}(\mathbf{I}-\hat{\mathbf{H}})\widehat{\mathrm{Var}(\mathbf{Y})}(\mathbf{I}-\hat{\mathbf{H}})'\hat{\mathbf{A}}^{-1}(1-2\hat{\boldsymbol{\pi}})$, $\widehat{\mathrm{E}(G_u)} = \hat{\boldsymbol{\pi}}'(1-\hat{\boldsymbol{\pi}})$, and $\widehat{\mathrm{Var}(G_u)} = (1-2\hat{\boldsymbol{\pi}})'(\mathbf{I}-\hat{\mathbf{H}})\widehat{\mathrm{Var}(\mathbf{Y})}(\mathbf{I}-\hat{\mathbf{H}})'(1-2\hat{\boldsymbol{\pi}})$, where $\hat{\mathbf{H}} = \hat{\mathbf{A}}\mathbf{Z}(\mathbf{Z}'\hat{\mathbf{A}}\hat{\mathbf{V}}^{-1}\hat{\mathbf{A}}\mathbf{Z})^{-1}\mathbf{Z}'\hat{\mathbf{A}}\hat{\mathbf{V}}^{-1}$ with $\mathbf{Z}' = (\mathbf{Z}_1', \cdots, \mathbf{Z}_n')$ being a $(J+p) \times (nTJ)$ matrix, $\hat{\mathbf{A}} = \mathrm{diag}(\hat{\pi}_{it}^{(j)}(1 - \hat{\pi}_{it}^{(j)}))$ and $\widehat{\mathrm{Var}(\mathbf{Y})} = \mathrm{diag}(\mathbf{V}_1(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}), \cdots, \mathbf{V}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}))$. Notice that the proposed statistics can be reduced to Pan's statistics [7] for binary responses when $J = 1$.

3 Simulation Study and Example

To evaluate the performance of the proposed tests for the proportional odds model fit in terms of type I error rate and the power, a model considered by Lin [9] is denoted by,

$$\mathrm{logit}(\eta_{it}^{(j)}) = \lambda_{1j} + 0.25\, D_{1i} + 1.5\, X_{1it} + \beta\, D_{1i} X_{1it},$$

where the monotone difference intercepts are assigned by $\lambda_1 = (-0.5, 0, 0.5)$, $D_{1i} = 0$ for $i \leq n/2$ and 1 otherwise, X_{1it} follows a uniform distribution on $(-1, 1)$ for $i = 1, \ldots, n$, $t = 1, 2, 3$, $j = 1, 2, 3$ and $n = (100, 500, 1000)$. The pairwise correlations between the observations at the three occasions within a subject are assumed to be 0.5. The covariance matrix of the uniform random variables $(X_{1i1}, X_{1i2}, X_{1i3})$ is set by

$$\boldsymbol{\Sigma} = \frac{1}{3} \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$

The correlated random numbers (u_1, u_2) based on $\boldsymbol{\Sigma}$ are generated by $u_1 \sim \mathrm{Unif}(-1, 1)$ and $u_2 = \mathbf{C}\, \Lambda^{1/2}\, u_1$, where $\mathbf{C} = (c_1, c_2, c_3)$ and $\Lambda = (a_1, a_2, a_3)$ are the eigenvectors and the eigenvalues of $\boldsymbol{\Sigma}$, respectively. The coefficient β is ranged from 0 to 5 to compute the empirical type I error rate ($\beta = 0$) and the powers ($\beta = 1, 2, 3, 4, 5$) of the proposed tests for detecting the interaction term. Table 1 shows the power comparison of the proposed tests and Lin's tests [9] utilizing four variants of 'working' correlation matrices: the independent, AR(1), exchangeable and unspecified structures for 1000 replications. It indicates that for all correlation structures the empirical type I error rates of the proposed tests G_p and G_u compared with their asymptotic normal distributions are quite close to the 5 percent level of significance. The powers of the proposed tests G_p and G_u are dramatically larger than those of the test proposed by Lin [9]. It reveals that the powers of the proposed tests based on the global odds ratio as a measure of association dominate over those tests using particular 'working' correlation matrices.

Table 1. The empirical type I error rates and powers of proposed tests G_p and G_u under four variant working correlation structures and global odds ratio (GOR) for $n = 100, 500, 1000$ at 5% level of significant.

Test		Independent			AR(1)			Exchangeable			Unspecified			GOR		
G_p	β	100	500	1000	100	500	1000	100	500	1000	100	500	1000	100	500	1000
	0	0.052	0.055	0.058	0.042	0.059	0.051	0.052	0.050	0.052	0.062	0.058	0.054	0.049	0.051	0.052
	1	0.396	0.512	0.598	0.315	0.345	0.381	0.352	0.396	0.452	0.213	0.265	0.304	0.798	0.821	0.860
	2	0.545	0.612	0.687	0.490	0.516	0.539	0.518	0.531	0.564	0.367	0.419	0.435	0.864	0.887	0.905
	3	0.699	0.764	0.803	0.607	0.626	0.644	0.620	0.667	0.701	0.557	0.573	0.598	0.894	0.910	0.934
	4	0.761	0.821	0.894	0.717	0.840	0.919	0.733	0.758	0.798	0.648	0.739	0.765	0.928	0.942	0.970
	5	0.836	0.897	0.945	0.799	0.906	0.939	0.788	0.824	0.865	0.734	0.786	0.806	0.945	0.973	0.990
G_u																
	0	0.050	0.061	0.053	0.047	0.053	0.058	0.041	0.049	0.051	0.044	0.049	0.045	0.048	0.047	0.050
	1	0.767	0.795	0.825	0.698	0.712	0.721	0.715	0.724	0.754	0.670	0.681	0.698	0.932	0.948	0.971
	2	0.869	0.887	0.906	0.823	0.871	0.901	0.838	0.842	0.857	0.797	0.825	0.849	0.959	0.980	0.994
	3	0.949	0.967	0.988	0.943	0.955	0.968	0.939	0.945	0.960	0.936	0.942	0.957	0.988	0.998	1.000
	4	0.987	0.994	1.000	0.981	0.996	0.999	0.986	0.996	0.999	0.978	0.992	0.999	1.000	1.000	1.000
	5	0.996	1.000	1.000	1.000	1.000	1.000	0.995	0.999	0.999	0.994	0.999	1.000	1.000	1.000	1.000

Table 2. The parameter estimates under the marginal model and the global odds ratio model, their standard errors and Wald statistics as well as p-values for marijuana study.

Marginal covariates	$\hat{\boldsymbol{\theta}}$	$S(\hat{\boldsymbol{\theta}})$	Wald statistic	p-value	Associate covariates	$\hat{\boldsymbol{\alpha}}$	$S(\hat{\boldsymbol{\alpha}})$	Wald statistic	p-value
λ_1	3.112	0.302	106.373	0.000	$\boldsymbol{\Delta}$	-5.514	0.259	452.013	0.000
λ_2	4.213	0.329	163.159	0.000	$\boldsymbol{\Delta_1}$	2.951	0.178	307.002	0.000
gender	-0.850	0.287	9.393	0.002	$\boldsymbol{\Delta_2}$	1.199	0.128	87.947	0.000
time	-0.493	0.059	69.676	0.000	time	-0.041	0.007	30.318	0.000

An example employed to illustrate the application of the proposed tests was a study of marijuana use for teenagers where 237 thirteen-year-old children who had used marijuana in the 1976$\sim$1980 five consecutive years. The longitudinal data set, originally from the U.S. National Youth Survey (Elliot *et al.*, [18]), includes a trichotomous response variable (1: never; 2: not more than once a month; 3: more than once a month), and two covariates: annual time and gender of children. The aim of this study was to investigate the effects of time and gender on marijuana use for teenagers from five annual waves. Vermunt and Hagenaars [19] considered an additive model to fit this longitudinal ordinal data, $\text{logit}(\eta_{it}^{(j)}) = \lambda_j + \beta_1\, t + \beta_2\, X_i$, where t is the annual wave for $t = 1, \ldots, 5$ and X_i is the gender variable. $X_i = 0$ for boys; $i = 1, \ldots, 117$ and $X_i = 1$ for girls; $i = 118, \ldots, 237$. The global odds ratio model is expressed as

$$log\Psi_{ijk}(s,t) = \Delta + \Delta_j + \Delta_k + \boldsymbol{\gamma}'|s-t|,$$

where we assumed $\Psi_{ijk}(s,t)$ depends on time and $\Delta_{jk} = 0$.

Table 2 summarizes the estimates of parameters, their standard errors and Wald statistics as well as p-values under the marginal model and the global odds ratio model. The effects of time and gender are both highly significant ($p < 0.01$). With the null additive model, the p-values of test statistics G_p and G_u are 0.924 and 0.673, respectively, leading to the adequacy of the model.

4 Conclusion and Discussion

The main purpose of this article is to propose two goodness-of-fit tests using the global odds ratio as a measure of association for modeling longitudinal ordinal data. For large samples, the proposed test statistics are approximated by normal distributions. The proposed tests have the largest powers among all current tests using various working correlation structures and sample sizes. It is because that the 'working' correlation parameters are subject to an uncertainty of definition referred by Crowder [20], and global odds ratio can be regarded as a suitable general structure for the association of data.

Alternative competing classes for modeling repeated ordinal data based on random-effect growth models can be referred to Williamson *et al.* [3], Heagerty and Zeger [5] and Vermunt and Hagenaars [19], which will be the ongoing research.

References

[1] K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika*, vol. 73, pp.13-22, 1986.

[2] N. E. Breslow and D. G. Clayton, Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, vol. 88, pp.9-25, 1993.

[3] J. M. Williamson, K. Kim and S. R. Lipsitz, Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, vol. 90, pp.1432-1437, 1995.

[4] J. M. Williamson, S. R. Lipsitz and K. Kim, GEECAT and GEEGOR: computer programs for the analysis of correlated categorical response data, *Computer Methods and Programs in Biomedicine*, vol. 58, pp.25-34, 1999.

[5] P. J. Heagerty and S. L. Zeger, Marginal regression models for clustered ordinal measurements, *Journal of the American Statistical Association*, vol. 91, pp.1024-1036, 1996.

[6] K. F. Yu and W. Yuan, Regression models for unbalanced longitudinal ordinal data: computer software and a simulation study, *Computer Methods and Programs in Biomedicine*, vol. 75, pp.195-200, 2004.

[7] W. Pan, Goodness-of-fit tests for GEE with correlated binary data, *Scandinavian Journal of Statistics*, vol. 29, pp.101-110, 2002.

[8] K. C. Lin, Y. J. Chen and Y. Shyr, A nonparametric smoothing method for assessing GEE models with longitudinal binary data, *Statistics in Medicine*, vol. 27, pp.4428-4439, 2008.

[9] K. C. Lin, Goodness-of-fit tests for modeling longitudinal ordinal data, *Computational Statistics and Data Analysis*, vol. 54, pp.1872-1880, 2010.

[10] P. McCullagh, Regression models for ordinal data, *Journal of the Royal Statistical Society*, vol. B42, pp.109-142, 1980.

[11] C. C. Clogg and E. S. Shihadeh, Statistical Models for Ordinal Data, Thousand Oaks, 1994.

[12] A. Agresti, Modelling ordered categorical data: recent advances and future challenges, *Statistics in Medicine*, vol. 18, pp.2191-2207, 1999.

[13] S. R. Lipsitz, K. Kim and L. Zhao, Analysis of repeated categorical data using generalized estimating equations, *Statistics in Medicine*, vol. 13, pp.1149-1163, 1994.

[14] R. L. Prentice, Correlated Binary Regression With Covariates Specific to Each Binary Observation, *Biometrics*, vol. 44, pp.1033-1048, 1988.

[15] S. R. Lipsitz, N. M. Laird and D. P. Harrington, Generalized Estimating Equations for Correlated Binary Data: Using the Odds Ratio as a Measure of Association, *Biometrika*, vol. 73, pp.153-160, 1991.

[16] R. L. Plackett, A Class of Bivariate Distribution, *Journal of the American Statistical Association*, vol. 60, pp.516-522, 1965.

[17] K. C. Lin, Y. J. Chen and C. Y. Liu, Goodness-of-fit tests for GEE models with longitudinal ordinal data using a global odds ratio, *ICIC Express Letters*, vol. 5, pp.1821-1826, 2011.

[18] D. S. Elliot, D. Huizinga and S. Menard, Multiple Problem Youth: Delinquence, Substance Use and Mental Health Problems, Springer-Verlag: New York, 1989.

[19] J. K. Vermunt and J. A. Hagenaars, Ordinal longitudinal data analysis. In: Hauspie R.C., Cameron N., Molinari L. (Ed.), Methods in Human Growth Research, Cambridge University Press, 2004.

[20] M. Crowder, On the use of a working correlation matrix in using generalized linear models for repeated measures, *Biometrika*, vol. 82, pp.407-410, 1995.